



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Order-parameter-aided temperature-accelerated sampling for the exploration of crystal polymorphism and solid-liquid phase transitions

T. Q. Yu, P. Y. Chen, M. Chen, A. Samanta, E.
Vanden-Eijnden, M. Tuckerman

May 7, 2014

Journal of Chemical Physics

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Order-parameter-aided temperature-accelerated sampling for the exploration of crystal polymorphism and solid-liquid phase transitions

Tang-Qing Yu*

*Courant Institute of Mathematical Sciences,
New York University, New York, NY 10012 , USA*

Pei-Yang Chen[†]

Department of Chemistry, New York University, New York, NY 10003 , USA

Ming Chen[†]

Department of Chemistry, New York University, New York, NY 10003 , USA

Amit Samanta[†]

*Program in Applied and Computational Mathematics,
Princeton University, Princeton, NJ 08544, USA and
Lawrence Livermore National Laboratory, Livermore, CA 94550, USA[†]*

Eric Vanden-Eijnden[†]

*Courant Institute of Mathematical Sciences,
New York University, New York, NY 10012 , USA*

Mark Tuckerman[§]

*Dept. of Chemistry New York University, New York, NY 10003 , USA
Courant Institute of Mathematical Sciences,
New York University, New York, NY 10012 , USA and
NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

(Dated: May 5, 2014)

Major/minor classification: Physical Sciences/Chemistry

ABSTRACT

The problem of predicting polymorphism in atomic and molecular crystals constitutes a significant challenge both experimentally and theoretically. From the theoretical viewpoint, polymorphism prediction falls into the general class of problems characterized by an underlying rough energy landscape, and consequently, free energy based enhanced sampling approaches can be brought to bear on the problem. In this paper, we build on a scheme previously introduced by two of the authors in which the lengths and angles of the supercell are targeted for enhanced sampling via temperature accelerated adiabatic free energy dynamics [T. Q. Yu and M. E. Tuckerman *Phys. Rev. Lett.* **107**, 015701 (2011)]. Here, that framework is expanded to include general order parameters that distinguish different crystalline arrangements as target collective variables for enhanced sampling. The resulting free energy surface, being of quite high dimension, is nontrivial to reconstruct, and we discuss one particular strategy for performing the free energy analysis. The method is applied to the study of polymorphism in xenon crystals at high pressure and temperature using the Steinhardt order parameters without and with the supercell included in the set of collective variables. The expected fcc and bcc structures are obtained, and when the supercell parameters are included as collective variables, we also find several new structures, including fcc states with hcp stacking faults. We also apply the new method to the solid-liquid phase transition in copper at 1300 K using the same Steinhardt order parameters. Our method is able to melt and refreeze the system repeatedly, and the free energy profile can be obtained with high efficiency.

I. INTRODUCTION

In the study of crystals, polymorphism refers to the ability of certain compounds to form multiple stable structures, a phenomenon that has significant ramifications for pharmaceuticals, high-energy materials, and organic electronics¹. Experimental determination of all of the relevant crystal structures of a particular compound under a given set of conditions is both lengthy and costly. Therefore, computational approaches for *a priori* polymorphism prediction, if sufficiently accurate and efficient, can potentially play an important role in the understanding and designing crystals in these and other fields.

Considerable effort has been invested over several decades in the prediction of crystal structures, and numerous theoretical methods have been developed². Despite notable successes in the computational prediction of crystal polymorphs^{3–5}, such predictions are far from routine, and the problem remains an important outstanding challenge. The most common approach starts with candidate structures obtained by packing the molecules according to the symmetry operations of the most common space groups, performing a local optimization, and then evaluating the associated lattice energy. Although this approach produces many candidate structures, it may produce false positives or miss solid forms that are not perfect crystals, including mixed structures and stable defects⁶, and it is not likely to identify structures that crystallize into rare space groups⁷. Moreover, this approach relies on a harmonic approximation to the calculation of lattice phonon frequencies of the candidate structures in order to determine thermodynamic properties such as the free energy. While the harmonic approximation is often sufficient for low-temperature crystals governed by strong intermolecular interactions, when systems are dominated by weak interactions, as is the case in many organic molecular crystals, anharmonic effects become important, and alternate approaches are needed. Obtaining fully anharmonic thermal contributions requires a methodology based directly on free energy generation. This category of techniques, however, entails the considerable challenge of sampling a complex and rough energy landscape in order to obtain the relative populations of the different polymorphs. Because of this, polymorphism prediction has been compared to the conformational exploration of proteins⁸. Although the two challenges are very different, they share important features, and consequently, some of the methods developed for biophysical structure prediction can be adapted for crystal

polymorphism exploration^{8–12}.

Following this idea, we recently introduced a new free energy approach for the discovery and thermodynamic ranking of polymorphs of atomic and molecular crystals^{6,13}. The technique is derived from the recognition that temperature can be employed to accelerate barrier crossing, an approach that has been shared by a variety of schemes including simulated annealing¹⁴, simulated tempering^{15,16}, parallel tempering^{17,18}, and temperature accelerated dynamics¹⁹. In our algorithm, we identify a set of collective variables (CVs) capable of distinguishing different states and subject them to a high-temperature heat bath. At the same time, these variables are also assigned high masses in order to effect an adiabatic decoupling of these variables from the remaining degrees of freedom. This approach is termed adiabatic free energy dynamics (AFED)^{20,21}. In the limit of perfect adiabatic decoupling, it can be proved that the CVs move on the correct potential of mean force surface, which is equivalent to the free energy surface²⁰. The approach of Refs. [6,13], which we call Crystal-AFED, is an adaptation of the AFED scheme to the isothermal-isobaric ensemble, in which the cell lengths and angles, or equivalently, the elements of the full cell matrix, are employed as the target CVs. Using these CVs, the polymorphs of crystalline benzene were studied⁶, and it was found that 500 ps of simulation time were sufficient to identify all of the stable polymorphs at 2 GPa and 100 K, and after just 5 ns, the free energy differences were sufficiently converged to resolve a controversy concerning the structure of the benzene II polymorph. This study also highlighted the importance of entropic contributions in the stabilization of the benzene II structure obtained.

Targeting the cell matrix alone for enhanced sampling in the discovery of crystal polymorphs is useful when different polymorphs are characterized by very different unit cell shapes, so that the lengths and angles of the cell are able to distinguish different structures. In some cases, however, these parameters alone are insufficient to induce transitions between different solid forms, for example, when collective flipping of molecules is required to effect such a transition. In addition, predicting crystal formation from amorphous or glassy states generally cannot be easily accomplished using just the cell matrix. In these examples, transitions between polymorphs and crystallization from disordered states are better described with the aid of structural order parameters. We will refer to these as “inter-

nal order parameters” in order to distinguish them from the cell matrix. The well known Steinhardt order parameters²² constitute one such example. Recently, novel approaches to the generation of general order parameters for molecular crystals have been introduced^{23,24}. Despite these intriguing advances, the problem of developing such general order parameters remains a significant challenge. Structural order parameters are often used in the study of phase transitions, nucleation, liquid crystals, and various other applications. Therefore, targeting molecular order parameters in combination with the cell matrix could potentially constitute a powerful approach to the discovery of crystal polymorphs based on enhanced sampling and free energy surface generation. However, because general order parameters are complicated functions of the primitive Cartesian coordinates of the atoms in a system, they cannot be easily treated within the AFED scheme, which requires that the CVs be explicit coordinates in the system or that they be made explicit via a transformation to a set of generalized coordinates that contains the CVs as a subset of the larger coordinate set. It was shown^{25,26}, however, that this problem could be circumvented by introducing an extended phase-space approach, in which the CVs of interest are harmonically bound to the coordinates in the extended space. These extended coordinates are then subject to a high temperature and are adiabatically decoupled from the remainder of the system. The approach is termed temperature-accelerated molecular dynamics (TAMD)²⁵ or driven adiabatic free energy dynamics (d-AFED)²⁶. Recently, a Monte Carlo version of this approach was also introduced²⁷, which fits the general framework of heterogeneous multiscale methods^{28,29}. In the limit of perfect adiabatic decoupling and infinitely stiff harmonic coupling, TAMD/d-AFED can be proved to generate the correct free energy surface as a function of the CVs^{25,26,30}.

In this paper, we further develop our previous temperature-accelerated sampling approach for the discovery of crystal polymorphs to include internal order parameters as additional CVs via the TAMD/d-AFED framework. We apply the approach to the case of solid xenon at high pressure. Under the conditions chosen, we find that hitherto unexpected polymorphs and fcc structures with hcp stacking faults are obtained. As a second example, we study the liquid-solid phase transition of copper, which illustrates that our new approach can also be effective in studying transitions between ordered and amorphous states.

II. METHODOLOGY

A. Equations of motion

Consider a system containing N atoms with positions $\mathbf{r}_1, \dots, \mathbf{r}_N$ interacting via a potential $U(\mathbf{r}_1, \dots, \mathbf{r}_N) \equiv U(\mathbf{r})$ in a supercell described by three vectors \mathbf{a} , \mathbf{b} , and \mathbf{c} such that the volume V of the cell is $V = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. These three vectors are collected in the columns of a matrix \mathbf{h} according to

$$\mathbf{h} = \begin{pmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{pmatrix}, \quad (1)$$

which is referred to as the *cell matrix*; its determinant gives the cell volume: $V = \det(\mathbf{h})$. Suppose we are interested in the free energy surface as a function of $n \ll 3N$ collective variables (CVs) $q_1(\mathbf{r}), \dots, q_n(\mathbf{r})$. Under isothermal-isobaric conditions with external pressure P , the relevant free energy is the Gibbs free energy $G(s_1, \dots, s_n)$, which is obtained from the marginal probability distribution for the CVs $q_1(\mathbf{r}), \dots, q_n(\mathbf{r})$ to have corresponding values s_1, \dots, s_n . This free energy is, therefore, given by

$$G(s_1, \dots, s_n) = -\frac{1}{\beta} \ln \left[\frac{1}{\Delta} \int d\mathbf{h} \frac{e^{-\beta P \det(\mathbf{h})}}{[\det(\mathbf{h})]^2} \times \int_{D(\mathbf{h})} d\mathbf{r}_1 \cdots d\mathbf{r}_N e^{-\beta U(\mathbf{r})} \prod_{\alpha=1}^n \delta(q_\alpha(\mathbf{r}) - s_\alpha) \right] \quad (2)$$

where $D(\mathbf{h})$ is the spatial domain defined by the cell matrix, $\beta = 1/k_B T$, $\Delta = \Delta(N, P, T)$ is the isothermal-isobaric partition function. The s variables are also known as *coarse-grained* variables (CGVs). We can also include the cell matrix in the set of CVs, in which case the relevant free energy is

$$G(s_1, \dots, s_n, \mathbf{h}) = -\frac{1}{\beta} \ln \left[\frac{e^{-\beta P \det(\mathbf{h})}}{\Delta [\det(\mathbf{h})]^2} \int_{D(\mathbf{h})} d\mathbf{r}_1 \cdots d\mathbf{r}_N e^{-\beta U(\mathbf{r})} \prod_{\alpha=1}^n \delta(q_\alpha(\mathbf{r}) - s_\alpha) \right] \quad (3)$$

In the Crystal-AFED approach of Refs. [6,13], the cell matrix alone was selected as the set of target CVs, and it was possible to drive these variables directly with a high temperature and adiabatic decoupling. However, when including internal order parameters in the set of CVs, this is no longer possible, and it is necessary to employ the extended phase-space

approach of Refs. [25,26]. In this case, we write the product of δ -functions $\prod_{\alpha} \delta(q_{\alpha}(\mathbf{r}) - s_{\alpha})$ as the limit of a product of Gaussian functions according to

$$\prod_{\alpha=1}^n \delta(q_{\alpha}(\mathbf{r}) - s_{\alpha}) = \prod_{\alpha=1}^n \lim_{\kappa_{\alpha} \rightarrow \infty} \left(\frac{\beta \kappa_{\alpha}}{2\pi} \right)^{1/2} \exp \left[-\frac{\beta \kappa_{\alpha}}{2} (q_{\alpha}(\mathbf{r}) - s_{\alpha})^2 \right] \quad (4)$$

In practice, the inverse width parameters κ_{α} are taken to be large but finite. Thus, if Eq. (4) is substituted into Eq. (2) or Eq. (3), the result is a modification of the potential $U(\mathbf{r})$ by the addition of a harmonic potential $V_{\{\kappa\}}(\mathbf{r}, s)$ given by

$$V_{\{\kappa\}}(\mathbf{r}, s) = \frac{1}{2} \sum_{\alpha=1}^n \kappa_{\alpha} (q_{\alpha}(\mathbf{r}) - s_{\alpha})^2 \quad (5)$$

where $s \equiv s_1, \dots, s_n$, so that the potential for the extended phase space becomes

$$V(\mathbf{r}, s) = U(\mathbf{r}) + V_{\{\kappa\}}(\mathbf{r}, s) \quad (6)$$

For finite κ_{α} , we obtain approximations $G_{\{\kappa\}}(s_1, \dots, s_n; \beta) \equiv G_{\{\kappa\}}(s; \beta)$ or $G_{\{\kappa\}}(s_1, \dots, s_n, \mathbf{h}; \beta) \equiv G_{\{\kappa\}}(s, \mathbf{h}; \beta)$ to the Gibbs free energy surfaces as functions either of s alone or of both s and \mathbf{h} , that approach the exact value as $\kappa_{\alpha} \rightarrow \infty$.

A set of temperature-accelerated equations of motion capable of generating the Gibbs surfaces in either Eq. (2) or (3) is based on the Martyna-Tobias-Klein (MTK) equations of motion^{31,32}, for which measure-preserving integrators have been developed^{33,34}. Introducing two temperatures T_s for the order parameters, and T_h for the cell matrix, the equations of motion take the form

$$\begin{aligned} \dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} + \frac{\mathbf{p}_g}{W} \mathbf{r}_i \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \frac{\mathbf{p}_g}{W} \mathbf{p}_i - \frac{1}{N_f} \frac{\text{Tr}[\mathbf{p}_g]}{W} \mathbf{p}_i + \text{Bath}(T) \\ \dot{\mathbf{h}} &= \frac{\mathbf{p}_g \mathbf{h}}{W} \\ \dot{\mathbf{p}}_g &= \det(\mathbf{h}) [\mathbf{P}^{(\text{int})} - P\mathbf{I}] + \frac{1}{N_f} \sum_{i=1}^N \frac{\mathbf{p}_i^2}{m_i} \mathbf{I} + \text{Bath}(T_h) \\ \dot{s}_{\alpha} &= \frac{p_{s_{\alpha}}}{\mu_{\alpha}} \\ \dot{p}_{s_{\alpha}} &= \kappa_{\alpha} (q_{\alpha}(\mathbf{r}) - s_{\alpha}) + \text{Bath}(T_s). \end{aligned} \quad (7)$$

Here, “Bath” refers to some heat bath coupling, e.g., Nosé-Hoover chains³⁵, generalized Gaussian Moment thermostats³⁶, a Langevin bath^{37–39}, ...

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i} - \sum_{\alpha} \kappa_{\alpha} (q_{\alpha}(\mathbf{r}) - s_{\alpha}) \frac{\partial q_{\alpha}}{\partial \mathbf{r}_i} + \mathbf{f}_i^{(\text{constr})} \quad (8)$$

is the total force on the i th atom, including forces from any holonomic constraints, and N_f is the effective number of degrees of freedom. The matrix \mathbf{p}_g is a 3×3 matrix that serves as a barostat to control the fluctuations of the pressure tensor estimator, which is given by

$$\mathbf{P}^{(\text{int})}(\mathbf{p}, \mathbf{r}) = \frac{1}{\det(\mathbf{h})} \sum_{i=1}^N \left[\frac{\mathbf{p}_i \otimes \mathbf{p}_i}{m_i} + \mathbf{F}_i \otimes \mathbf{r}_i \right]. \quad (9)$$

The mass-like parameters W and μ_α that determines the time scales for the motion of the cell matrix and internal order parameters, respectively, are determined by the conditions $W = k_B T_h \tau_h^2$, $\mu_\alpha = k_B T_s \tau_s^2$, where τ_h and τ_s time scales relevant to the motion of each set of CVs, respectively. The units of the harmonic coupling parameters κ_α depend on the choice of the associated CVs, so that the harmonic coupling term in Eq. 5 has units of an energy. Finally, we note that there are two choices for the parameter T_h in Eqs. (7). If the Gibbs free energy in Eq. (2) is sought, then T_h should be set equal to the physical temperature T , while if the free energy in Eq. (3) is the goal, then T_h should be set equal to T_s . In Ref. 13, we developed the algorithms needed to integrate the original equations of Crystal-AFED⁶ when holonomic constraints are imposed on a system, and the introduction of the extended variables in Eqs. (7) requires no change to this procedure. Consider, first, the case when $T_h = T$, so that the Gibbs free energy surface in Eq. (2) is sought. When the extended variables s are sufficiently slow, they are driven by forces generated by averaging over the motion of the remaining variables. In this case, the motion of s is effectively governed by

$$\begin{aligned} \mu_\alpha \ddot{s}_\alpha &= -\frac{\partial}{\partial s_\alpha} \langle V_{\{\kappa\}}(\mathbf{r}, s) \rangle_{s+\text{Bath}(T_s)} \\ &= \langle \kappa_\alpha (q_\alpha(\mathbf{r}) - s_\alpha) \rangle_s + \text{Bath}(T_s) \end{aligned} \quad (10)$$

where $\langle \cdots \rangle_s$ indicates an ensemble average over the degrees of freedom $(\mathbf{r}, \mathbf{p}, \mathbf{h}, \mathbf{p}_g)$ at fixed

values of $s = (s_1, \dots, s_n)$. Thus, the ensemble average of $\langle \kappa_\alpha (q_\alpha(\mathbf{r}) - s_\alpha) \rangle_s$ can be written as

$$\begin{aligned}
\langle \kappa_\alpha (q_\alpha(\mathbf{r}) - s_\alpha) \rangle_s &= \frac{C}{\Delta_{\{\kappa\}}(s, \beta)} \int d^N \mathbf{p} \int d\mathbf{p}_g \int \frac{d\mathbf{h}}{[\det(\mathbf{h})]^2} \int_{D(\mathbf{h})} d^N \mathbf{r} [\kappa_\alpha (q_\alpha(\mathbf{r}) - s_\alpha)] \\
&\quad \times \exp \left\{ -\beta [H(\mathbf{p}, \mathbf{r}) + \text{Tr}[\mathbf{p}_g^T \mathbf{p}_g]/2W + P \det(\mathbf{h}) + V_{\{\kappa\}}(\mathbf{r}, s)] \right\} \\
&= \beta^{-1} \frac{\partial}{\partial s_\alpha} \ln \Delta_{\{\kappa\}}(s; \beta) \\
&= -\frac{\partial G_{\{\kappa\}}(s; \beta)}{\partial s_\alpha}. \tag{11}
\end{aligned}$$

Here $H(\mathbf{p}, \mathbf{r})$ is the physical Hamiltonian $H(\mathbf{p}, \mathbf{r}) = \sum_i \mathbf{p}_i^2/2m_i + U(\mathbf{r})$, C is a normalization constant, and $G_{\{\kappa\}}(s; \beta)$ is the finite- κ approximation to the true Gibbs potential of mean force surface, which is also the free energy surface. $\Delta_{\{\kappa\}}(s; \beta)$ is the finite- κ partition function at fixed s_1, \dots, s_n :

$$\begin{aligned}
\Delta_{\{\kappa\}}(s; \beta) &= C \int d^N \mathbf{p} \int d\mathbf{p}_g \int \frac{d\mathbf{h}}{[\det(\mathbf{h})]^2} \int_{D(\mathbf{h})} d^N \mathbf{r} \\
&\quad \times \exp \left\{ -\beta \left[H(\mathbf{p}, \mathbf{r}) + \text{Tr}[\mathbf{p}_g^T \mathbf{p}_g]/2W + P \det(\mathbf{h}) + V_{\{\kappa\}}(\mathbf{r}, s) \right] \right\}. \tag{12}
\end{aligned}$$

Since Eq. (10) generates the density $P_{\{\kappa\}}(s; \beta_s, \beta) \propto \exp(-\beta_s G_{\{\kappa\}}(s; \beta))$, the Gibbs free energy surface can be constructed from a normalized histogram $\tilde{P}_{\{\kappa\}}^{(\text{adb})}(s; \beta, \beta_s)$ collected during the integration of Eqs. (7) under adiabatic ('adb') conditions using

$$G_{\{\kappa\}}(s; \beta) = -\beta_s^{-1} \ln \tilde{P}_{\{\kappa\}}^{(\text{adb})}(s; \beta_s, \beta) + C', \tag{13}$$

According to Eq. (4), in the limit $\{\kappa_\alpha \rightarrow \infty\}$, $G_{\{\kappa\}}(s; \beta)$ converges to the exact Gibbs free energy surface $G(s)$ in Eq. 2. In practice, $\{\kappa_\alpha\}$ are chosen sufficiently large that the difference between $G_{\{\kappa\}}$ and G is negligible, noting that the error is bounded and of order $\max_\alpha (1/\kappa_\alpha)^{30}$. When $T_{\mathbf{h}} = T_s$, a similar argument can be made for the full Gibbs free energy surface in Eq. (3).

B. FES construction and representation

While a free energy surface can be obtained directly by accumulating a histogram over the course of a simulation, this approach has the obvious limitation that the number of bins rapidly increases with the dimensionality of the surface. In order to address this problem, we make use of the single-sweep method proposed in Ref. 40. In this approach, TAMD/d-AFED is first used to explore the configuration space only, and centers are deposited in the space of the target CVs. Next, free energy gradients (i.e. mean forces) calculated at each center are used to construct the FES using a set of radial basis functions (RBF) as interpolants. The use of RBFs as interpolants in the construction of an FES is a technique that can be applied on sparse grids with regular or irregular boundaries, and therefore a high-dimensional generalization of the thermodynamic integration method⁴¹

In order to apply the single-sweep method, we first represent the free energy surface $G_{\{\kappa\}}(s; \beta)$ in Eq. 13 as a linear combination, denoted $\tilde{G}(z)$, of RBFs. This expansion takes the form

$$\tilde{G}(z) = \sum_{k=1}^K a_k \phi_{\sigma}(|z - z^{(k)}|) \quad (14)$$

Here z is the full set of coarse-grained variables, *i.e.*, $z = (s_1, \dots, s_n)$ if Eq. (2) is sought, or it is the full set of coarse-grained variables and the box matrix, *i.e.*, $z = (s_1, \dots, s_n, \mathbf{h})$ if Eq. (3) is sought. Note that each element of \mathbf{h} , which is generally taken to be upper or lower triangular when off-diagonal elements are needed, is treated as an independent CV when constructing the linear combination in Eq. (14). One possible choice of $\phi_{\sigma}(r)$ is a Gaussian kernel of width σ : $\phi_{\sigma}(r) = \exp(-r^2/2\sigma^2)$. The K Gaussian centers, $z^{(k)}$, are chosen along a CV trajectory generated from a TAMD/d-AFED simulation and the optimal coefficients a_k and σ are determined via minimization of the the cost function

$$E(a, \sigma) = \sum_{k=1}^K |f^{(k)} + \nabla_{z^{(k)}} \tilde{G}(z)|^2, \quad (15)$$

where the numerical mean force $f^{(k)}$ is obtained from restrained MD simulations performed at the locations of the centers $z^{(k)}$. The mean force $f_{\alpha}^{(k)}$ on the coarse-grained variable s_{α} is given by the average

$$f_{\alpha}^{(k)} = \langle \kappa_{\alpha}(q_{\alpha}(\mathbf{r}) - s_{\alpha}) \rangle_{z^{(k)}} \quad (16)$$

where z is fixed at $z^{(k)}$ in the NPT simulation. The mean force on $h_{\mu\nu}$, when \mathbf{h} is included in the FES, is calculated from

$$f_{h_{\mu\nu}} = \det(\mathbf{h}) \cdot \sum_{\gamma} [h_{\mu\gamma}^{-1} \langle (P_{\gamma\nu}^{(\text{int})} - P\delta_{\gamma\nu}) \rangle_{z^{(k)}}] \quad (17)$$

with both the coarse-grained variables s_1, \dots, s_n and \mathbf{h} held fixed at the selected centers $z^{(k)}$, which now include centers for the box matrix. Note that, when \mathbf{h} is held fixed, the ensemble is equivalent to an $N\mathbf{h}T$ ensemble. For a fixed σ value, minimization of the cost function in Eq. (15) leads to linear equations⁴⁰, and the coefficients a_k can be obtained using any linear solver. The calculation can be carried out for a range of σ values in order to find an optimal choice for σ that leads to the lowest overall error⁴⁰. Note that one can also use the modified cost function of Monteferrante *et al.*⁴².

The functional form of the free energy surface via the linear combination of radial basis functions in Eq. (14) may not provide a transparent picture of the free energy landscape, especially when the dimension is high. In order to locate all of the minima on the free energy surface thus constructed, we run many optimizations in the CGV space using, for example, a steepest-descent algorithm, from an ensemble of initial points. Since the centers used in the reconstruction cover the important regions of the configuration space, they serve well as the initial points. After these minima are found, the string method^{43,44} can be used to locate saddle points on the high-dimensional free energy surface as well as the minimum free energy paths (MFEPs). Once this information is available, a network can be generated in which the minima are the vertices/nodes and any two minima are connected by an edge if there is a MFEP connecting them directly. We also assign a weight to each edge, which is the free energy of the saddle on the MFEP. In this way, we obtain a weighted graph representation for the FES, similar to a recent scheme introduced for protein free energy landscapes⁴⁵. Such a network is a reduced representation of the surface that captures its most salient features, including critical points and their associated free energy values. Any further analysis, particularly analysis involving integration over the CVs, can then be performed using Eq. (14).

III. ORDER PARAMETERS USED IN THIS STUDY

The Steinhardt order parameters²², $Q_l, l = 3, 4, 6, \dots$, are widely used to identify crystal structures. Q_4 and Q_6 can distinguish the simple cubic (sc), face-centred cubic (fcc), body-centred cubic (bcc), and hexagonal closest packing (hcp) crystal structures found in the solid phases of spherical particles. For molecular crystals, one typically requires more general order parameters^{23,24,46}. Here, we employ the continuous version of the general Steinhardt order parameter given in Ref. [47,48]. Thus, the global order parameters used as CVs in our study take the form

$$Q_l = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^l |Q_{lm}|^2 \right]^{1/2}, \quad (18)$$

where

$$Q_{lm} = \frac{1}{NN_{\text{coor}}} \sum_{b=1}^{N_b} f_c(r_b) Y_{lm}(\hat{\mathbf{r}}_b). \quad (19)$$

Here, N_b is the total number of atom pairs separated by a distance r_{max} , N is the total number of atoms, and N_{coor} is the first-shell coordination number of each atom. Y_{lm} are the spherical harmonics, $\hat{\mathbf{r}}_b$ is the unit vector along the direction \mathbf{r}_b , i.e., $\hat{\mathbf{r}}_b = (\sin \theta_b \cos \phi_b, \sin \theta_b \sin \phi_b, \cos \theta_b)$. While the normalization factor in Eq. 19 should be N_b , the quantity N_b varies as the phase changes, which will cause discontinuities in the function Q_l . For this reason, we choose NN_{coor} as the normalization factor for convenience. In this study, $N_{\text{coor}} = 6$. $f_c(r)$ is a smooth switching function defined as

$$f_c(r) = \begin{cases} 1, & r < r_{\text{min}} \\ \frac{1}{2} \left\{ \cos \left[\frac{(r-r_{\text{min}})}{r_{\text{max}}-r_{\text{min}}} \pi \right] + 1 \right\}, & r_{\text{min}} < r \leq r_{\text{max}} \\ 0, & r > r_{\text{max}} \end{cases}$$

This function is used to remove discontinuities in the original definition of the Steinhardt parameters²² that occur when each bond is switched off at a specific radius r_{max} . The value of r_{max} is determined from the end of the first peak in the radial distribution function.

In order to describe the bond orientation for each atom, we employ the local Steinhardt

order parameters, $q_l, l = 3, 4, 6, \dots$ defined as

$$q_l(i) = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^l |q_{lm}(i)|^2 \right]^{1/2}, \quad (20)$$

where

$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\hat{\mathbf{r}}_{ij}). \quad (21)$$

Here, $N_b(i)$ is the number of all pairs connecting to atom i within a cutoff r_{\max} as above. The parameter $q_l(i)$ can be used to distinguish different atomic neighbor environments, i.e., a bcc arrangement will have a very different q_l value from an fcc arrangement. Another set of local Steinhardt order parameters, denoted as $w_l(i)$ and defined as

$$w_l(i) = \frac{\sum_{m_1+m_2+m_3=0} \begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} q_{lm_1}(i) q_{lm_2}(i) q_{lm_3}(i)}{\left(\sum_{m=-l}^l |q_{lm}(i)|^2 \right)^{3/2}}, \quad (22)$$

is employed to analyze solid structures. Since the temperature in our studies is close to the melting point, thermal fluctuations are very large. Consequently, the distribution of these local order parameters is rather broad, which diminishes their ability to distinguish different crystal structures. Therefore, averaged versions of these bond order parameters, introduced previously by Lechner and Dellago⁴⁹, which are more sensitive to different crystal structures, are also employed in the present study, although other local order parameters are also possible⁵⁰. The averaged bond order parameters $\bar{q}_l(i)$ and $\bar{w}_l(i)$ have the same definitions as Eq. (20) and Eq. (22) except that q_{lm} is replaced by \bar{q}_{lm} , which is given by

$$\bar{q}_{lm}(i) = \frac{1}{\tilde{N}_b(i)} \sum_{k=0}^{\tilde{N}_b(i)} q_{lm}(k), \quad (23)$$

where summation runs over all neighbors of particle i , including particle i , itself. It is obvious that averaged bond order parameters account for the second shell.

IV. XENON POLYMORPHISM AT HIGH PRESSURE AND MELTING POINT

Recent simulation studies^{51,52} of solid xenon shows that its crystal structure undergoes a fcc-bcc transition at high pressures (25 GPa \sim 30 GPa) close to melting point (2700 K \sim 2900

K)^{51,52}. These studies also show that bcc grows naturally from an fcc-liquid mixed-phase state, and the authors reported a phase diagram with the fcc-bcc-liquid triple point near 25 GPa and 2700 K. Subsequent theoretical investigations challenged the exact location of this triple point^{53,54}. Enhanced sampling techniques have the distinct advantage over studies of this type in that they allow the metastability and coexistence behavior to be investigated on the basis of the free energy surface. To our knowledge, no such calculations have been performed for this problem, which makes it an interesting test case for our new approach.

A. Implementation

We consider a system containing 4,000 Xenon atoms initially in a $10 \times 10 \times 10$ fcc arrangement. The \mathbf{h} matrix was constrained to be orthorhombic, and therefore only the diagonal elements (the cell lengths a , b and c) were used. We chose a temperature of 2700 K and a pressure of 25 GPa as the imposed external conditions of the simulation. Interactions were described by a Buckingham potential⁵⁵, which is believed to improve on the simpler Lennard-Jones potential for condensed systems of noble gas atoms⁵⁶. The accuracy of this model has been verified against experiment⁵². All the simulations were performed using the PINY_MD code.^{57,58}

We first studied the FES with the variables Q_6 and Q_4 as the target CVs. The extended variables were maintained at 1.5×10^5 K for enhanced sampling with $\tau_s = 50$ ps, $\tau_{\mathbf{h}} = 0.5$ ps, and $r_{\max} = 4.5$ Å. The coupling constant κ was the same for Q_6 and Q_4 and set to 1×10^{10} K. In a second study, we investigated how the FES changes when \mathbf{h} is added to Q_6 and Q_4 in the CV set, which leads to a five-dimensional FES. The temperature of the extended variables and \mathbf{h} matrix was maintained at 1×10^5 K for enhanced sampling with $\tau_s = 50$ ps, $\tau_{\mathbf{h}} = 1$ ps, and $r_{\max} = 4.5$ Å. The coupling constant κ for Q_6 and Q_4 were again taken to be 1×10^{10} K. Finally, in order to prevent the system from sampling liquid or glassy states, we restricted $Q_6 > 0.7$.

In both cases we used the following protocol in the simulations: Ten 5 ns long trajectories starting from the fcc crystal were generated in order to explore the configuration space by integrating Eqs. (7) with a time step of 5 fs. To reconstruct the FES associated with Q_4 and Q_6 alone, 764 centers were deposited along the trajectories in such a way that any new

center should be at a distance of at least 0.01 from any previously deposited center. Within a neighborhood of 0.003 of each center, all instantaneous forces on the corresponding extended variables were collected from the sampling run for the calculation the mean forces at that center. In order to reconstruct the FES associated with Q_4 , Q_6 , and \mathbf{h} , the cell lengths were first scaled by a factor of 1/30 so that the range of cell lengths were comparable to those of Q_6 and Q_4 . Then, 2462 centers were deposited in the five-dimensional CV space. The mean forces at those centers were calculated from restrained MD simulations of 10 ps each. Because of the scaling of \mathbf{h} , the mean forces needed to be scaled up by a factor of 30 in the corresponding components. Finally, the stability of all the structures corresponding to minima on the FES were tested using a standard isothermal-isobaric simulation of length 5 ps with a fully-flexible box.

B. Results and discussion

Figures 1(a) and 1(b) show the trajectories of Q_4 and of the cell lengths when the enhanced sampling targeted only the Q_4 and Q_6 variables. Both trajectories show two stable phases, namely fcc and bcc, under the conditions of the simulation. The two-dimensional FES subsequently calculated by the single-sweep method is shown in Fig. 2. The basins corresponding to the fcc and bcc structures can clearly be seen on this surface, which also shows the minimum free-energy path between these structures generated using the string method (note that we neglected the tensor M entering the definition of the MFEP in this calculation³⁰ since our primary goal was to calculate the free energies of the minima and the saddle point on the landscape). Fig. 3 shows the free energy profile along the MFEP, from which we can estimate the free energy difference between bcc and fcc structure as 2.5 meV/atom (or 58 cal/mol). Note that for a homogeneous 4,000-atom system, *e.g.*, a pure fcc or pure bcc structure, finite-size effects are small, so that the microscopic unit meV/atom can be transformed directly into a macroscopic unit, *e.g.*, kcal/mol. The very small FE difference indicates that the bcc and fcc are thermodynamically equally stable at 2700 K and 25 GPa. Our results support the previous two-phase studies^{51,52} in that the bcc structure is stable at 2700 K and 25 GPa, and therefore, the fcc-bcc-liquid triple point should be close to this condition in the P - T plane. Table I gives the energy decomposition

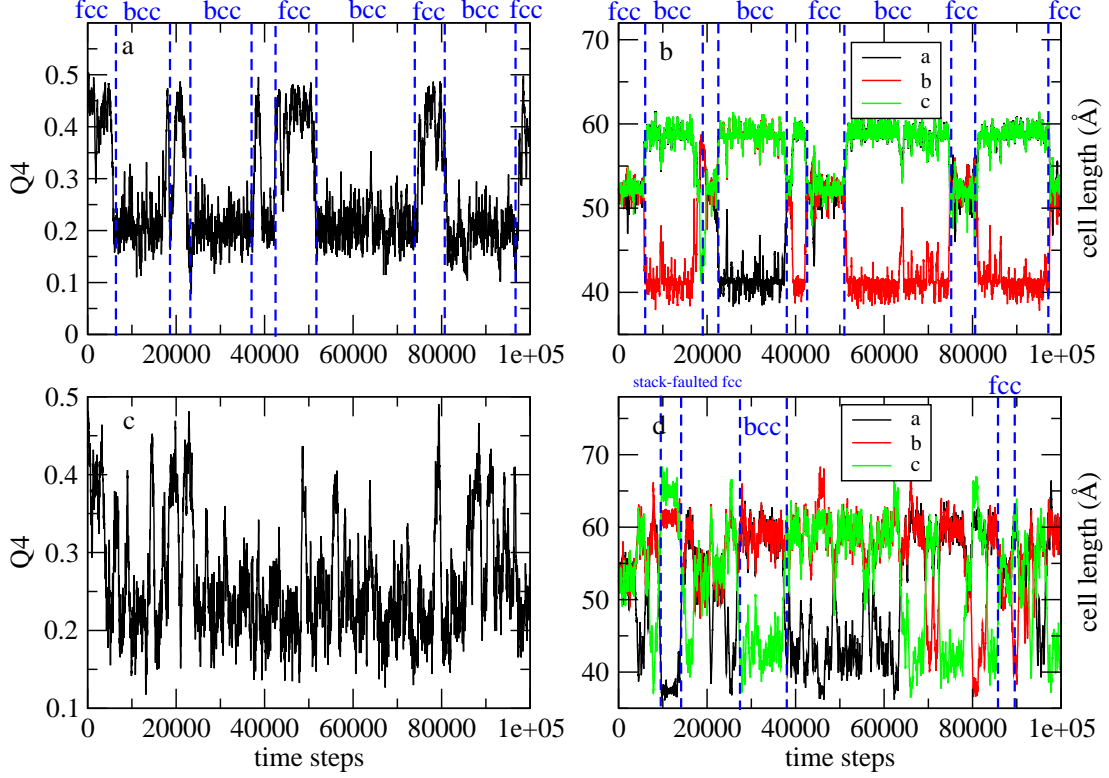


FIG. 1: Panels a and c: Trajectory of Q_4 . Panels b and d: Trajectory of cell lengths. Panels a and b are from temperature-accelerated sampling with only Q_4 and Q_6 as CVs, while in panels c and d, Q_4 , Q_6 and the cell lengths are used as CVs. $Q_4 > 0.3$ corresponds to the fcc structure; $Q_4 < 0.3$ corresponds to the bcc structure.

analysis for the fcc and bcc structures. Not unexpectedly perhaps, we find that the bcc structure is stabilized by entropic effects. Such stabilization was also predicted in relatively early studies based on simple Lennard-Jones models⁵⁹. The high entropy of the bcc structure dominates over the enthalpy, thus allowing it to become more stable than the fcc structure. As the temperature decreases and entropic effects become less important, we predict that fcc becomes the dominant state, which is well known for Xenon and was also seen in early studies⁵⁹. This is a good example showing that the free energy is the right thermodynamic state function for predicting crystal polymorphism, particularly when the temperature is high.

When the three cell lengths are added to the set of CVs, a larger set of metastable structures is explored at 2700 K and 25 GPa. Figures 1(c) and 1(d) show the resulting trajectory. A full five-dimensional FES was constructed using the single-sweep method with the mean forces on the CVs as input. Free energy minima were also located on this surface.

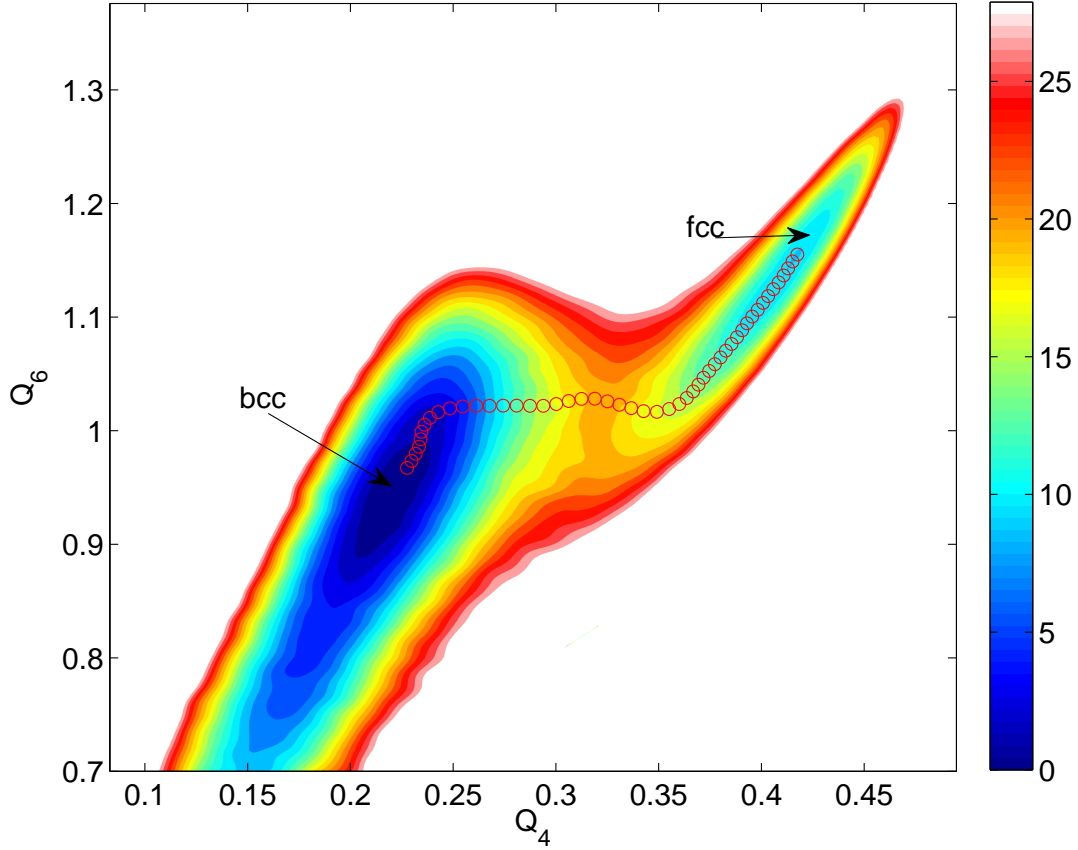


FIG. 2: Free energy surface as a function of Q_4 and Q_6 when these variables are used as the sole target CVs. The red circles represent the minimum energy path connecting the bcc and fcc basins, as calculated from the string method based on the constructed FES. The free energy values, in eV, correspond to the total free energy of the system.

TABLE I: Free energy decomposition analysis for the fcc and bcc structures. ΔG is the free energy difference; ΔH is the enthalpy difference; ΔV is the volume difference; ΔS is the entropy difference, $\Delta G = \Delta H - T\Delta S$. All energies are in meV/atom. Enthalpy and volume are calculated from a 10,000-step MD simulation. $T = 2700$ K and $P_{\text{ext}} = 25$ GPa.

	ΔG	ΔH	ΔE	$P_{\text{ext}}\Delta V$	$T\Delta S$
bcc - fcc	-2.5	12.8	10.8	2.0	15.2

The locations of the centers used in the FES construction (grey dots) and of the minima (colored filled circles) are shown in Fig. 4 after projections in the space of Q_4 and Q_6 , and in Fig. 5 in the space of the three cell lengths. The fact that the minima are tightly clustered in the projection on the Q_4 - Q_6 plane (see Fig. 4) but are well separated in the space of

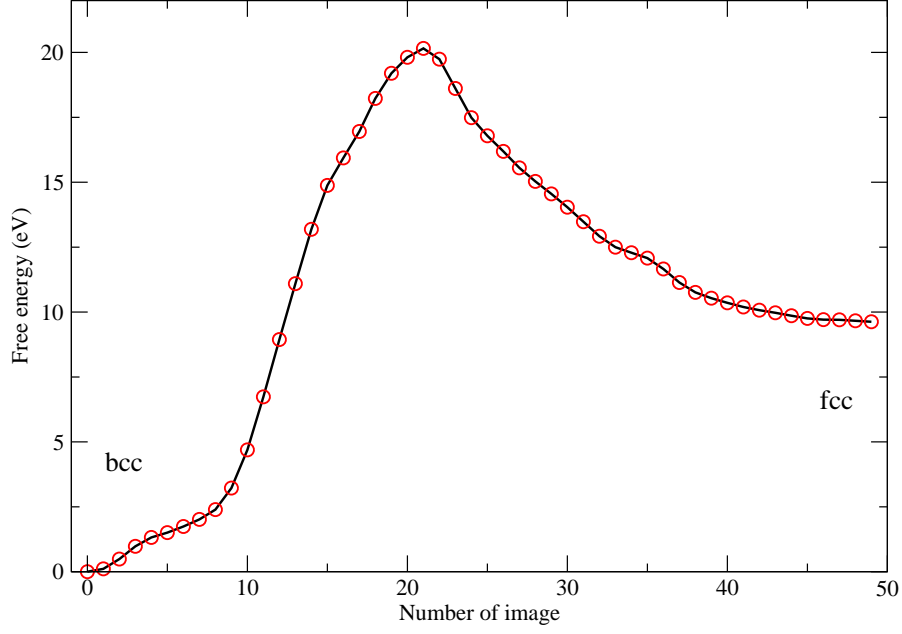


FIG. 3: Free energy profile along the minimum free energy path at 2700 K and 25 GPa.

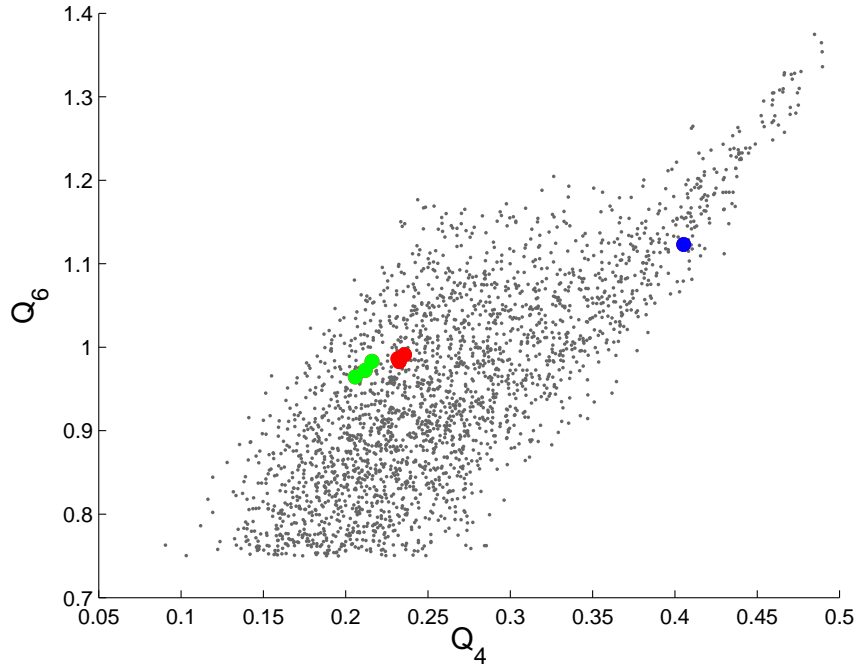


FIG. 4: Distribution of centers and minima in the space of Q_4 and Q_6 . Grey dots are centers deposited in the space of the collective variables. The red filled circles are minima for bcc, the blue is minima for fcc and the green are minima for new metastable states identified as fcc with stacking fault later.

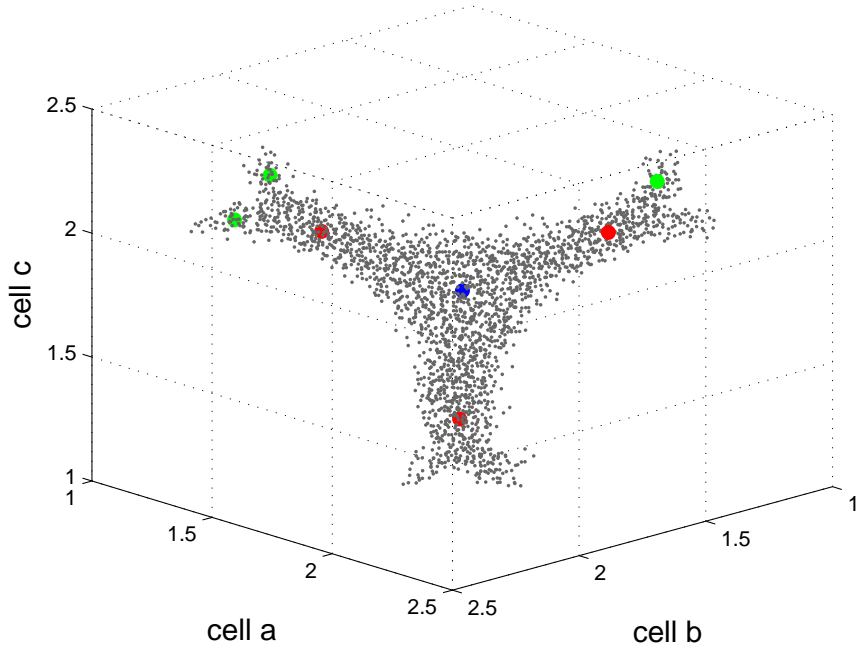


FIG. 5: Distribution of centers and minima in the space of cell lengths. The cell lengths have been scaled down by a factor of $1/30$ in units of Å. Grey dots are centers deposited in the space of the collective variables. The red filled circles are minima for bcc, the blue is minima for fcc and the green are minima for new metastable states identified as fcc with stacking fault later.

cell lengths (see Fig. 5) shows that using Q_4 and Q_6 alone as CVs is insufficient to identify more metastable solid forms for this system; indeed, the location of these minima cannot be accurately resolved in the lower dimensional Q_4 - Q_6 space. Thus, we see that the inclusion of the cell matrix *in conjunction with the internal order parameters* is essential. Finally, in order to better visualize the five-dimensional FES, we created a network representation of it as described in methodology (see Fig. 6). Fig. 6 also contains the free energy values (relative to the bcc structure) of the minima and saddle points. In order to avoid confusion, we should point out that the saddle or MFEP we obtained may not be the true solid-solid transition path, as finite-size effects⁶⁰ could lead to an overestimation of the free energy barriers.

The structures shown in the network diagram were further analyzed using local order parameters and averaged local order parameters. The values of the local and averaged local order parameters for all of the structures are shown in Fig. 7. In addition to the fcc and bcc structures already identified, we also find fcc crystals with stacking faults characterized by

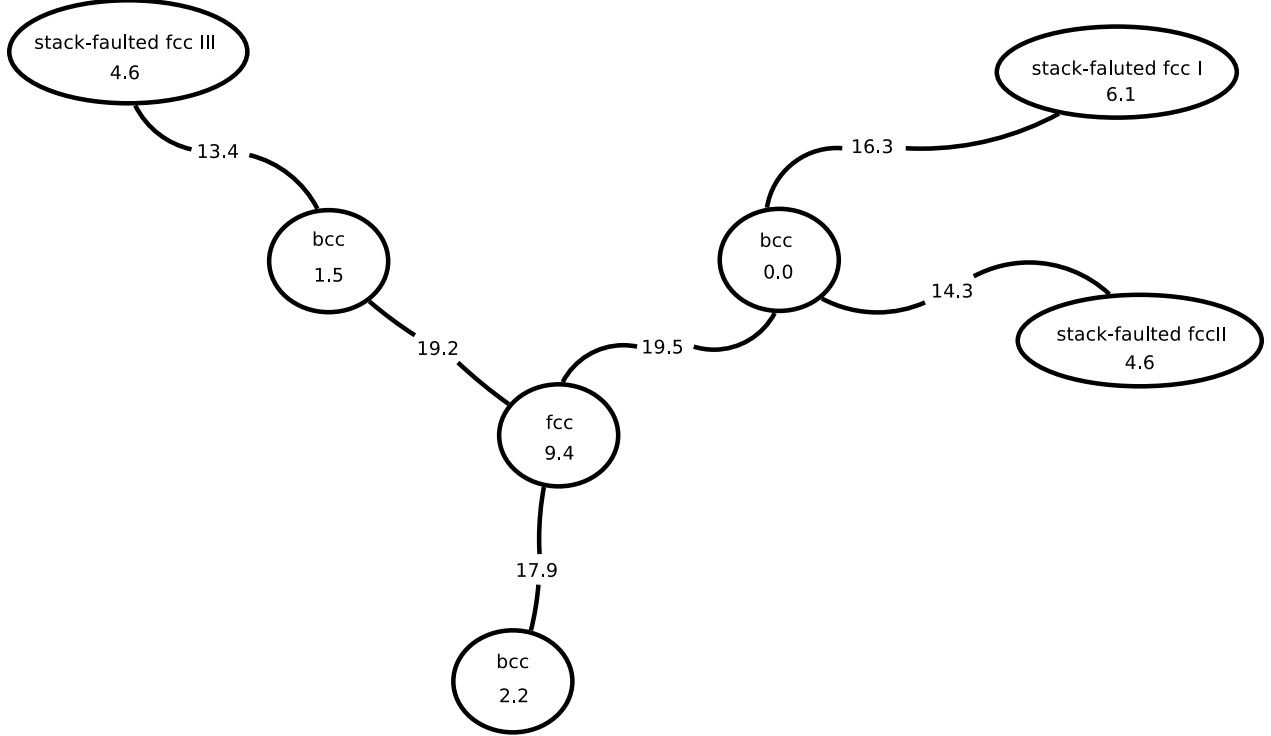


FIG. 6: Network representation of 5-dimensional free energy surface along Q_4 , Q_6 and cell lengths. The numbers assigned to each node correspond to free energies at the minima; the numbers assigned to the edges are the free energy values for the corresponding saddle points. The free energy is in eV.

an hcp structure. Even though fcc-hcp polytypism is expected to exist, it is still impressive that our sampling can locate these structures. Fig. 7 (a) and (b) show that there exist two types of atoms within such a structure. One type is characterized by order parameters located in the fcc region (for example, $w_4 \in [-0.10, -0.06]$ and $q_4 \in [0.17, 0.19]$ in (b)), and the other type is characterized by order parameters located in the hcp region (for example, $w_4 \in [-0.1, 0.02]$ and $q_4 \in [0.13, 0.15]$ in (b)). (See also Ref. [49] for the order parameter values that characterize the hcp structure.) Fig. 8 shows snapshots of several such structures with atoms colored based on their (w_4, q_4) values. These pictures clearly show the fcc and hcp arrangement occurring in an alternating pattern in these structures. Even though we sample just a few fcc structures with such stacking faults, we expect that longer runs or runs employing a higher temperature for \mathbf{h} and the extended variables could yield even more of these. Experimental evidence for the existence of such states has been reported in pressure-induced Martensitic fcc-to-hcp transformations⁶¹, where they are interpreted as intermediate states between fcc and hcp. In order to see whether the defects in our fcc structure are

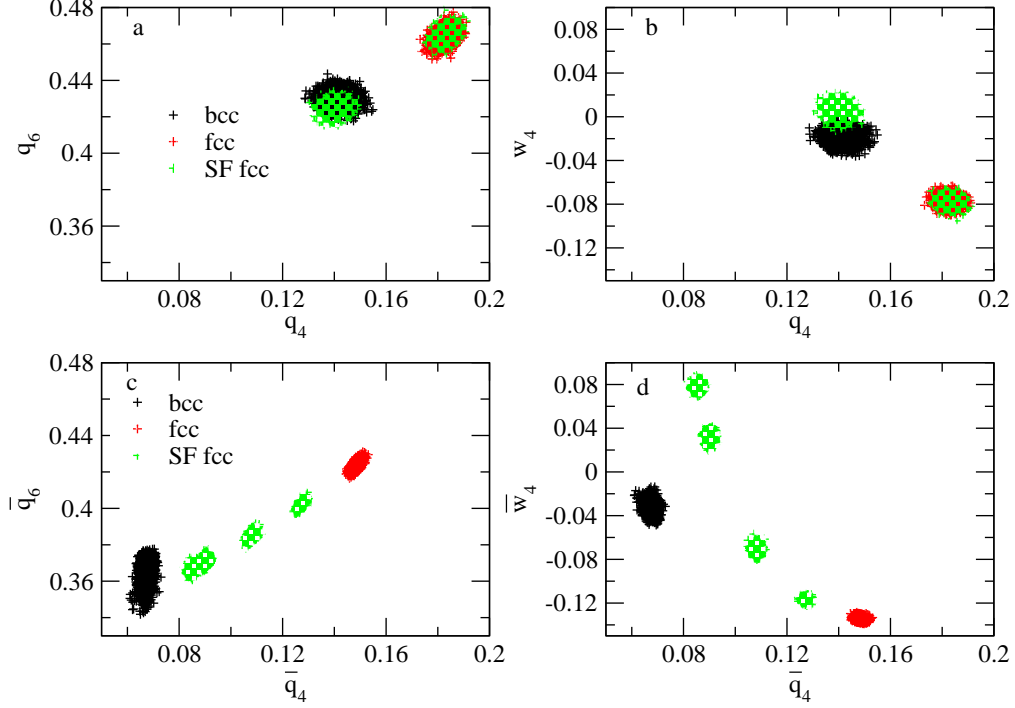


FIG. 7: Comparison of the distributions of local order parameters for several solid forms. The variables q_6 , q_4 and w_4 are the local bond order parameters; \bar{q}_6 , \bar{q}_4 and \bar{w}_4 are the averaged bond order parameters. In order to minimize thermal smearing, order parameters for each atom are obtained by averaging over 200 equilibrated configurations. 4,000 dots corresponding to 4,000 atoms are given for each solid form. The stack-faulted fcc structures (SF fcc in legend) show two types of atoms according to their local order parameters (see green clusters in a and b). The averaged bond order parameters include the second neighbor shell. Therefore, the atoms of the stack-faulted fcc split into more types as the local environment of an atom becomes more diversified (see green clusters in c and d).

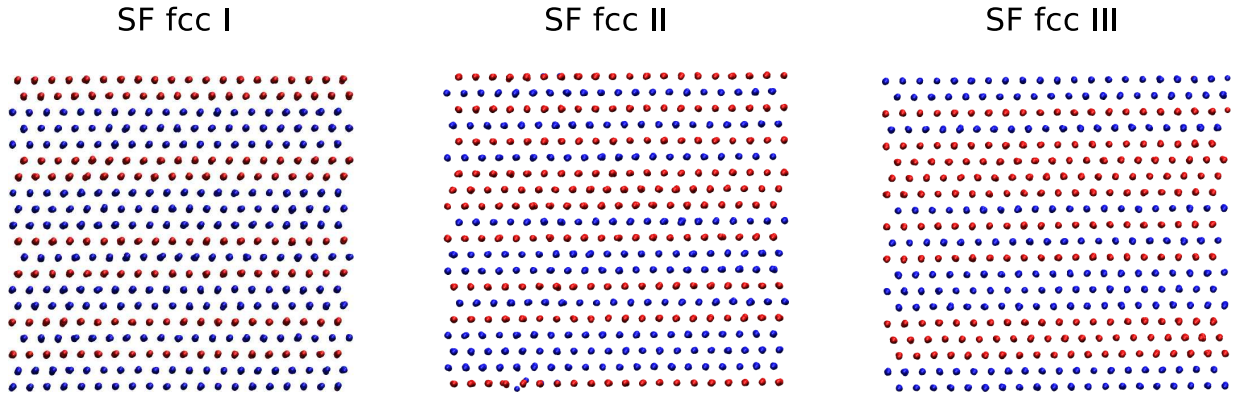


FIG. 8: Pictures of the fcc with stacking fault. Red corresponds to fcc atom type, with $w_4 \in [-0.10, -0.06]$ and $q_4 \in [0.17, 0.19]$; blue corresponds to hcp atom type, with $w_4 \in [-0.1, 0.02]$ and $q_4 \in [0.13, 0.15]$. SF fcc means fcc with stacking fault.

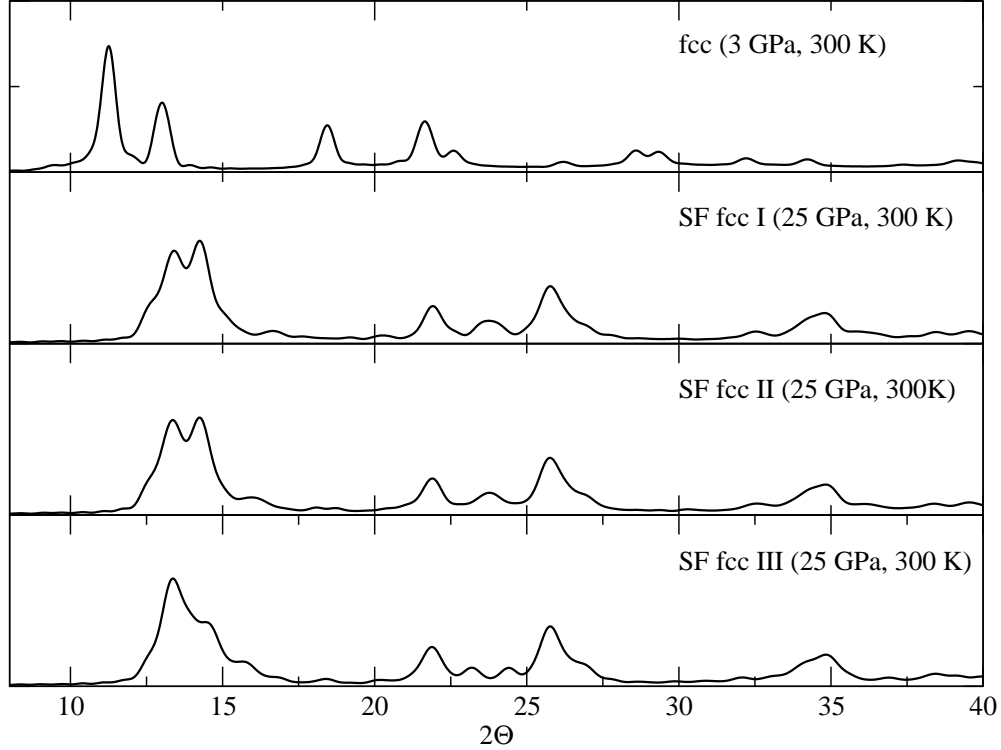


FIG. 9: Calculated powder diffraction pattern of the fcc structures with stacking fault. The Debye package⁶² was used to obtain the powder patterns, and the calculation is based on the Debye scattering equation⁶³. The temperature and pressure are taken to be the experimental conditions for direct comparison. The wavelength is taken to be 0.7 \AA so that the powder diffraction pattern of the fcc structure at 3 GPa is identical to the experimental one in Fig. 1 of Reference 61. SF fcc means fcc with stacking fault.

indeed these intermediate states, we have calculated the powder diffraction pattern using the Debye package⁶². Fig. 9 shows the powder patterns of the three fcc structures with stacking faults. These match well the patterns shown in Fig. 1 of Reference 61. The bulk system is treated as a cluster in the calculation (no periodic boundary conditions) and therefore, finite-size effects may explain the tiny difference between the calculated and experimental powder diffraction patterns. From the comparison of the powder diffraction patterns, there is compelling evidence to suggest that the stacking-fault structures we obtained could resemble the intermediate states visited in the pressure-induced fcc-to-hcp transition.

Body centered orthorhombic (bco) and body centered tetragonal (bct) structures also appeared in our sampling. The radial distribution functions (RDFs) of the bco, bct, fcc, and bcc forms at 25 GPa and 2700 K are presented in Fig. 11 in order to show the structural difference. The RDFs for the fcc and bcc structures show different features, which can also

serve as an identifier for fcc and bcc, as was done in previous studies^{51,52}. The differences between bcc and bco/bct are sufficiently small that the RDFs are very similar. Therefore, in order to identify the bco/bct structure, we calculated their unit cell parameters. Unit cells at high temperature are generally difficult to determine. Since atom switching or atomic jump diffusion may happen in the bulk system at high temperatures, the trajectory-averaged position of an atom gives misleading information about the final lattice as many atoms can be off the lattice sites. Thus, in order to obtain reliable unit-cell information, we employed the following procedure: From four well-equilibrated configurations, we take each atom in each configuration as a “central” atom and find its neighboring atoms within the first peak of the RDF. For the bcc structure, the first peak of the RDF consists of both first and second shell atoms while for the fcc structure, it contains only first-shell atoms. The central atom together, with its neighbors, forms one unit, and we have 16,000 such units (each configuration having 4,000 atoms). Overlaying these units with the central atom at the same location and aligning them generates one central atom surrounded by several well-separated clusters. We then use a *K*-means clustering algorithm to find the clusters and calculate their centroids as the average positions of the atoms associated with the cluster. The centroid is the most probable position of the neighboring atoms. We, therefore, infer the unit cell information from these centroids. In order to illustrate the approach, Fig. 10 shows the clusters and their centroids thus obtained for the bcc and fcc structures. In addition, we have provided animations in the Supporting Information (SI) showing the clusters and centroids from different perspectives. The unit cell parameters of fcc, bcc and bco/bct are summarized in Table II. In addition to providing a reliable identification of the bco and bct structures, the high-temperature unit cell analysis we employed has an advantage over an analysis of the RDFs: Due to high-temperature thermal smearing, the bcc structure acquires a putative coordination number of fourteen if that coordination number is computed from an integration of the RDF up to the first minimum. However, the coordination number obtained from the high-temperature unit-cell analysis is eight as expected for the bcc structure.

The bct structure remains stable during an isothermal-isobaric (NPT) MD simulation with a fully flexible cell. The stability of the bco structure can only be verified when the simulation box is constrained to be orthorhombic. When we relax it in an NPT simulation

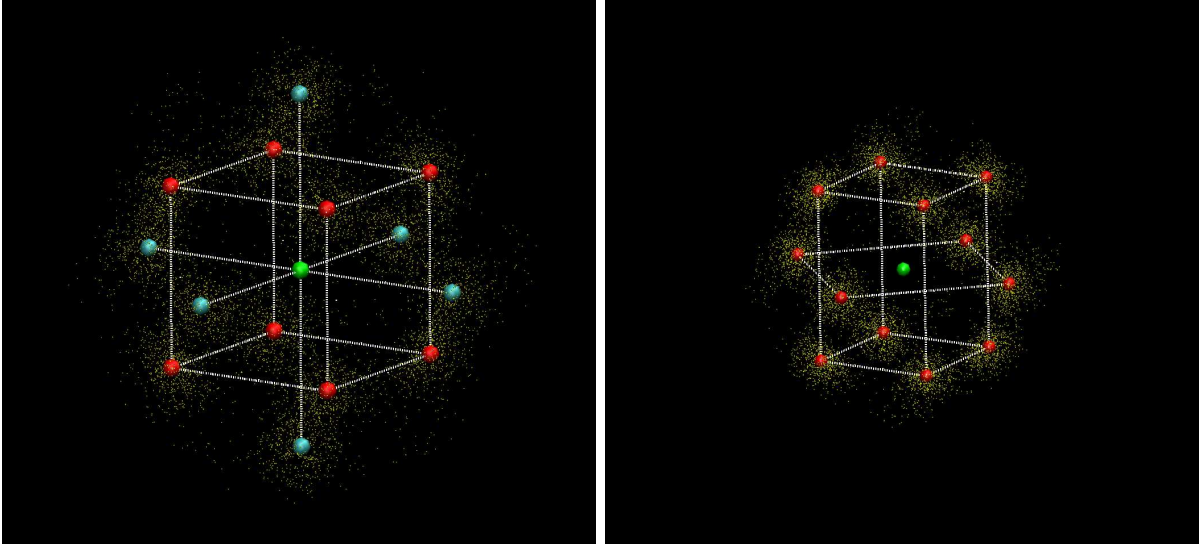


FIG. 10: Left panel: The bcc structure obtained by the K -means clustering approach (see text). The snapshot shows the central atom is shown in green, the red spheres represent the centroids of the clusters corresponding to the first-shell atoms, while the cyan spheres show the centroids of the clusters corresponding to the second-shell atoms. The small yellow spheres represent 8,000 randomly selected cluster points from 224,000 total points. Right panel: The fcc structure obtained by the K -means clustering approach. As with the bcc structure, the green sphere is the central atom, the red spheres are the centroids of the clusters corresponding to the first-shell atoms, and the small yellow spheres represent 8,000 randomly selected cluster points from 192,000 total points.

TABLE II: Unit cell parameters for fcc, bcc, and the predicted crystal forms bco and bct at 25 GPa and 2700 K. Cell lengths are in Å. $\alpha = \gamma = \beta = 90^\circ$.

	a	b	c	Z
fcc	5.27	5.27	5.27	4
bcc	4.25	4.25	4.25	2
bco	4.40	4.23	4.14	2
bct	4.22	4.27	4.27	2

with a fully flexible cell, the orthorhombic box shape tilts with final box angles being $\alpha = 85.4$, $\beta = 86.8$ and $\gamma = 91.0$, which implies that bco may not be a true stable/metastable state. This is reasonable as the FES on the sub-manifold of the orthorhombic cell constraint can have minima that may not persist when the constraint is removed. However, the success we have shown in locating this minimum, nevertheless, shows the power of the present approach in generating the FES and locating new crystal forms. Because bct/bco structures are only visited once in our trajectories and since their cell matrices have one short length

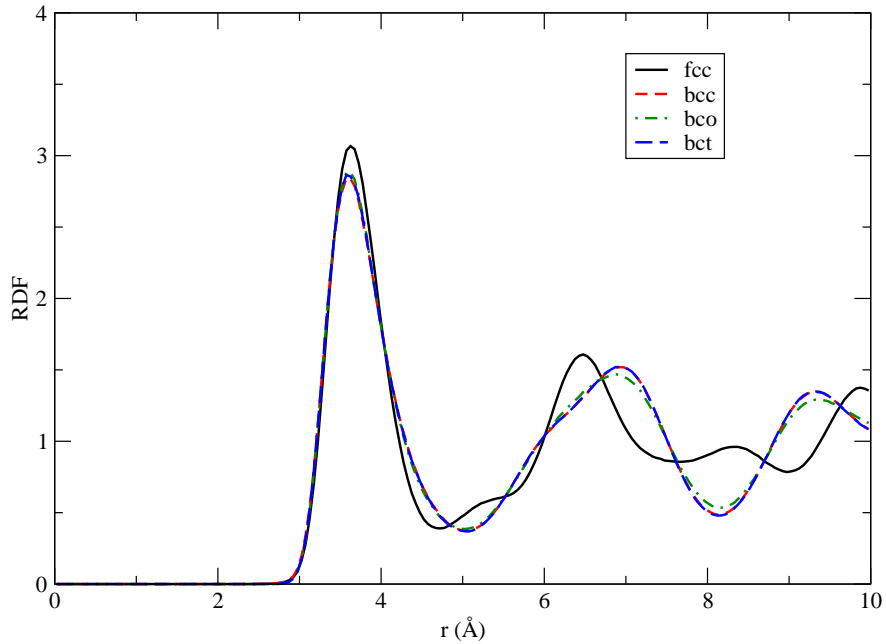


FIG. 11: Radial distribution function for several solid forms at 25 GPa and 2700 K.

of 36.5 Å, finite-size effects may play a role, and we, therefore, exclude them from the 5-dimensional FES construction, leaving them for future studies.

V. SOLID-LIQUID TRANSITIONS OF COPPER

Molecular dynamics (MD) simulations of melting at superheated conditions have provided atomistic insights into some of the theoretical models of the melting mechanism^{64,65}. Relying on MD simulations to observe the equilibrium melting of a solid is not optimal because, near the melting point, the melting transition is a rare event with a mean first passage time many orders of magnitude greater than characteristic lattice vibrational periods. Enhanced sampling methods, such as umbrella sampling and metadynamics, have been used to calculate free energy changes in solid-liquid transitions of ductile metals⁶⁶ and of ice/water^{47,67}. Our temperature-accelerated sampling method offers a robust way to calculate free energy changes for melting/freezing processes. Here, we use copper as an example to show that order-parameter-aided TAMD/d-AFED can render such a calculation very efficient.

A. Simulation details

The interatomic interactions were modeled using an Embedded Atom Method (EAM) potential for copper developed by Mishin et al.⁶⁸. Simulations were performed using a cell containing 4,000 copper atoms ($10 \times 10 \times 10$ unit cell). We have used the Steinhardt order parameters Q_6 and Q_4 as the collective variables to obtain the FES. In order to explore the free energy surface the initial system is slowly heated to the target temperature 1300 K under a pressure of 1 atm in a series of NPT simulations. Starting from the final structure of this first phase, we launched a 1.5 ns TAMD/d-AFED trajectory with the extended variables at 1×10^5 K. From this trajectory, we selected 15 configurations every 100 ps and used them as the initial configurations for 15 independent TAMD/d-AFED samplings of 1 ns each. The extended variables were maintained at 1×10^7 K for the purpose of enhanced sampling with $\tau_s = 7.7$ ps for Q_6 and $\tau_s = 15.5$ ps for Q_4 . We have taken 5×10^8 K and 1×10^8 K as the coupling constants for Q_4 and Q_6 , respectively. $\tau_h = 0.75$ ps. $r_{\max} = 2.75$ Å. Along these trajectories, a total of 933 centers were deposited at distances no less than 0.05 from each other. The mean force at a center was evaluated from the instantaneous forces sampled from TAMD/d-AFED trajectories when the corresponding extended variables were within 0.01 to that center. All the simulations were performed using the PINY_MD code^{57,58} with an integration time-step 1 fs.

B. Results and discussion

Along a representative TAMD/d-AFED sampling trajectory, we observed several melting and refreezing transitions. A portion of this trajectory has been rendered into a movie and is available in the SI. Fig. 12 shows the free energy surface in the variables Q_6 and Q_4 at 1300 K and 1 atm pressure, which is close to experimental melting point of 1360 K. The minimum free energy path (red circles in Fig. 12) on the resulting FES was calculated from the string method, and the free energy along it is shown in Fig. 13.

Different methods for estimating the melting temperature have been suggested in the literature, including the Z-method⁶⁹ and two-phase melting⁷⁰ to mention a few. One estimate of the melting temperature has been reported for the same EAM potential⁷¹. In this study,

a value of $T_m = 1300 \pm 15$ K was obtained using the superheating-supercooling hysteresis approach⁷², while T_m was found to be 1350 ± 20 K with the two-phase method. Finally, an average of these two values, 1325 K, was used as the melting point in this study. The small free energy difference of the solid and liquid state from our calculation (~ 4 eV for 4000 atoms or 1 meV per atom at 1300 K) is consistent with the previous predictions of the melting point.

We observe that the solid basin is populated with a host of metastable states that are characterized by different defects (mainly, vacancy-interstitial pairs, dislocations, interstitial clusters). In order to verify independently the existence of these metastable states on the high dimensional potential energy surface, we have used the configurations from TAMD/d-AFED trajectories to perform isothermal-isobaric MD relaxations for 50 ps. A majority of these relaxed configurations, possibly corresponding to metastable states, lie inside the solid basin. These locally stable states inside the solid basin correspond to point defects - different concentrations of vacancy interstitial pairs, defect clusters, etc. and line defects such as dislocations. The presence of multiple metastable states might suggest the existence of multiple melting pathways (*i.e.* the system can escape from the solid basin along different pathways), and this possibility will be the subject of a future mechanistic study. We further note that the MFEP obtained and the barrier calculated from the FES pertain to a relatively small system of 4,000 copper-atoms. It is, therefore, necessary to verify that this system size is sufficiently large to capture the critical nucleus under the conditions studied. Moreover, a histogram test^{73,74} should be performed in order to validate the CVs chosen. These validations lie beyond the scope of the present paper, whose focus is on the new methodology, however, they will be performed in future study focused specifically on the melting process. Nevertheless, the present example demonstrates, we believe, the ability of the new methodology to sample efficiently both the solid and liquid states.

VI. CONCLUSION

In this paper, we have shown that temperature accelerated techniques employing both the cell matrix and general order parameters as target collective variables lead to enhanced sampling both of crystal polymorphs and of solid-liquid phase transitions. Our scheme

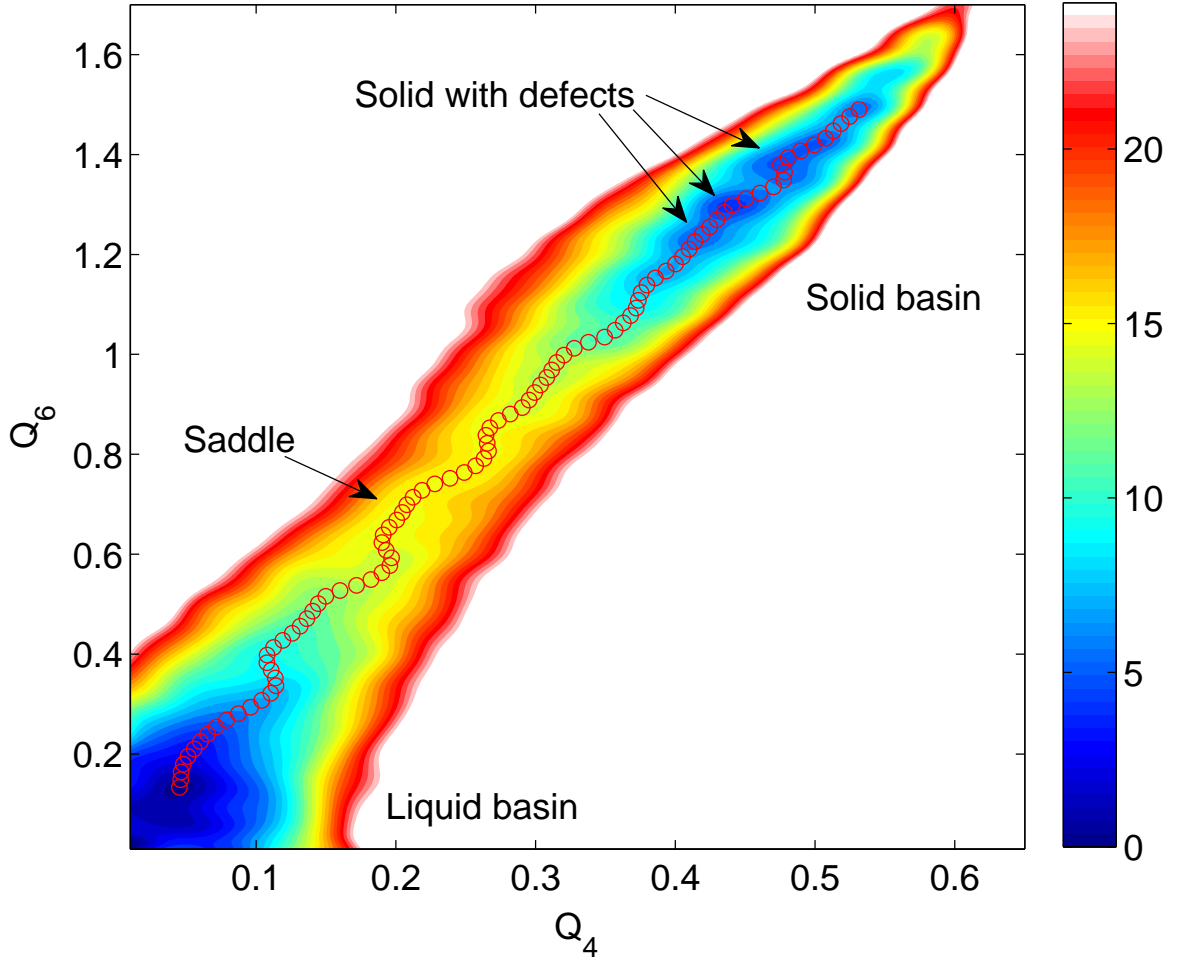


FIG. 12: The free energy surface in the $Q_6 - Q_4$ plane at 1300 K and 1 atm. The MFEP (red circles) is calculated from the string method based on the constructed free energy surface. The unit for free energy is eV.

is based on a combination of the crystal-AFED (adiabatic free energy dynamics) approach previously introduced by Yu and Tuckerman^{6,13} and the temperature accelerated MD/driven AFED^{25,26} adapted for the isothermal-isobaric ensemble and applied to the aforementioned order parameters. Since the resulting free energy surfaces are of relatively high dimension, we have discussed a robust approach for analyzing these surfaces, including the extraction of free energy values at basins/saddles from the string method and the representation of the FES as a network graph.

The method was applied in studies of polymorphism in xenon crystals at high pressure and temperature using the Steinhardt order parameters as collective variables and to the

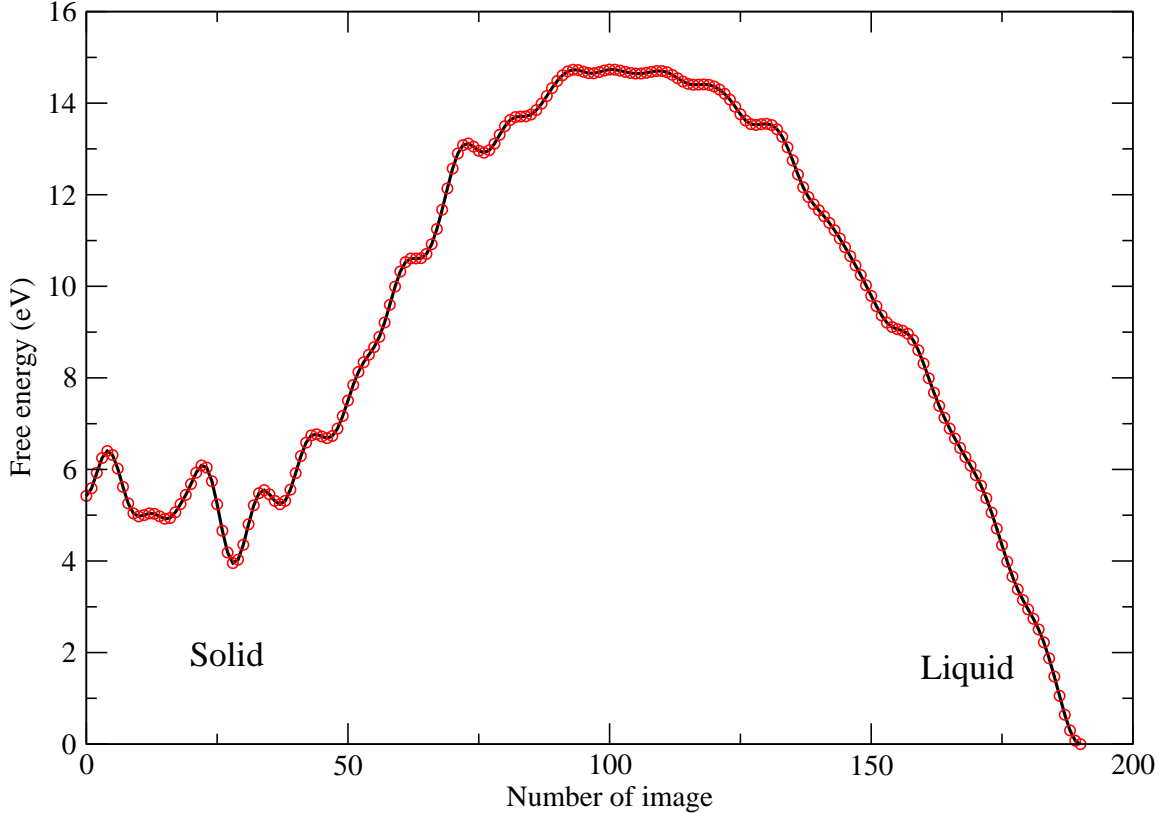


FIG. 13: The free energy profile along MFEP in $Q_6 - Q_4$ space at 1300 K, 1 atm.

solid-liquid transition in copper at 1300 K. In the xenon crystal, the expected fcc and bcc structures were recovered, and several additional structures, including an fcc state with a stacking fault, were identified. For copper, we showed that the enhanced sampling approach allows the free energy surface of the solid-liquid transition to be efficiently generated from the trajectories generated.

In implementing the new approach introduced, a number of conditions need to be considered. In general, we require the system size and box shape (orthorhombic, monoclinic, triclinic,...) to accommodate any possible unit cell size and shape of the crystal under study, and the search range for the method can be constrained to restrict the search to a particular class of structures of particular interest. In an isotropic system, such as the xenon example presented, restricting the box matrix to be diagonal, corresponding to an orthorhombic unit cell, proves sufficient to explore its polymorphs. For an anisotropic system, such as a molecular crystal, where molecular orientation is an important parameter, use of a fully flexible box is generally required, as proved to be the case in our recent study of crystalline

benzene⁶. Given that system size $N = N_a \times N_b \times N_c \times Z$, where N_a , N_b , and N_c are the numbers of replicas in each crystalline direction and Z is the number of asymmetric units in the unit cell of the crystal whence the sampling/search originates, we must choose N_a , N_b , and N_c such that N is a common multiple of any possible Z value. In many cases involving small organic molecules, the most common Z values are 2 and 4 but rare numbers must be considered if a crystal can support an uncommon unit cell shape or space group⁷. At the same time, N should be sufficiently large that structures with stable defects, such as were identified for crystalline benzene⁶, can be accounted for.

We expect that the framework outlined in this paper to be a potentially powerful approach for the discovery of different polymorphs in many crystalline systems and for inducing order/disorder transitions. In addition, such an approach could also serve as a tool to supplement machine learning techniques²⁴ in order to accelerate and aid in the process of fitting general order parameters for more complex molecular crystals, and this will constitute future work in this area.

Acknowledgments

The authors are grateful to David Chandler for useful discussions. A. S. was supported by the U.S. Department of Energy (DOE) at the Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. M. C. acknowledges the Margaret and Herman Sokol Doctoral Fellowship in the Sciences and support from the National Science Foundation (grant no. NSF CHE-1301314). M. E. T. acknowledges support from the National Science Foundation (grant Nos. NSF CHE-1012545 and NSF CHE-1301314). The research of E.V.-E. was supported in part by NSF Grant DMS07-08140 and ONR Grant N00014-11-1-0345. The research of T.-Q. Y. was supported in part by NIH Grant R01GM100472.

-
- * Electronic address: `tangqing.yu@nyu.edu`
- † These authors contributed equally to this work.
- ‡ Electronic address: `eve2@cims.nyu.edu`
- § Electronic address: `mark.tuckerman@nyu.edu`
- ¹ J. BERNSTEIN, *Polymorphism in Molecular Crystals*, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, 2002.
- ² S. M. WOODLEY and R. CATLOW, *Nature Materials* **7**, 937 (2008).
- ³ H. NOWELL, C. S. FRAMPTON, J. WAITE, and S. L. PRICE, *Acta Cryst.* **62**, 642 (2006).
- ⁴ R. PODESZWA, B. M. RICE, and K. SZALEWICZ, *Phys. Rev. Lett.* **101**, 115503 (2008).
- ⁵ A. B. KAZANTSEV, P. G. KARAMERTZANIS, C. S. ADJMAN, C. C. PANTELIDES, S. L. PRICE, P. T. A. GALEK, G. M. DAY, and A. J. CRUZ-CABEZA, *Intl. J. Pharmaceuticals* **418**, 168 (2011).
- ⁶ T.-Q. YU and M. E. TUCKERMAN, *Phys. Rev. Lett.* **107**, 015701 (2011).
- ⁷ T. GELBRICH, D. ROSSI, and U. J. GRIESSER, *Acta Cryst.* **E68**, o235 (2012).
- ⁸ J. D. DUNITZ and H. A. SCHERAGA, *Proc. Natl. Acad. Sci.* **101**, 14309 (2004).
- ⁹ R. MARTOŇÁK, A. LAIO, and M. PARRINELLO, *Phys. Rev. Lett.* **90**, 075503 (2003).
- ¹⁰ P. RAITERI, R. MARTOŇÁK, and M. PARRINELLO, *Angew. Chem. Int. Ed.* **44**, 3769 (2005).
- ¹¹ T. ZYKOVA-TIMAN, P. RAITERI, and M. PARRINELLO, *J. Phys. Chem. B.* **112**, 13231 (2008).
- ¹² P. G. KARAMERTZANIS, P. RAITERI, M. PARRINELLO, M. LESLIE, and S. L. PRICE, *J. Phys. Chem. B* **112**, 4298 (2008).
- ¹³ T.-Q. YU and M. E. TUCKERMAN, *Eur. Phys. J. Special Topics* **200**, 183 (2011).
- ¹⁴ S. KIRKPATRICK, C. D. GELATT, and M. P. VECCHI, *Science* **220**, 671 (1983).
- ¹⁵ E. MARINARI and G. PARISI, *Europhys. Lett.* **19**, 451 (1992).
- ¹⁶ P. N. VORONTSOV-VELYAMINOV and A. P. LYUBARTSEV, *Mol. Simulat.* **9**, 285 (1992).
- ¹⁷ R. H. SWENDSEN and J. S. WANG, *Phys. Rev. Lett.* **57**, 2607 (1986).
- ¹⁸ Y. SUGITA and Y. OKAMOTO, *Chem. Phys. Lett.* **314**, 141 (1999).
- ¹⁹ M. R. SORENSEN and A. F. VOTER, *J. Chem. Phys.* **112**, 9599 (2006).
- ²⁰ L. ROSSO, P. MINARY, Z. W. ZHU, and M. E. TUCKERMAN, *J. Chem. Phys.* **116**, 4389 (2002).

- ²¹ L. ROSSO and M. E. TUCKERMAN, *Mol. Simulat.* **28**, 91 (2002).
- ²² P. J. STEINHARDT, D. R. NELSON, and M. RONCHETTI, *Phys. Rev. B* **28**, 784 (1983).
- ²³ E. E. SANTISO and T. B. L., *J. Chem. Phys.* **134**, 064109 (2011).
- ²⁴ P. GEIGER and C. DELLAGO, *J. Chem. Phys.* **139**, 164105 (2013).
- ²⁵ L. MARAGLIANO and E. VANDEN-EIJNDEN, *Chem. Phys. Lett.* **426**, 168 (2006).
- ²⁶ J. B. ABRAMS and M. E. TUCKERMAN, *J. Phys. Chem. B* **112**, 15742 (2008).
- ²⁷ G. CICCOTTI and S. MELONI, *Phys. Chem. Chem. Phys.* **13**, 5952 (2011).
- ²⁸ W. E and B. ENGQUIST, *Commun. Math. Sci.* **1**, 87 (2003).
- ²⁹ E. WEINAN, B. ENGQUIST, X. LI, W. REN, and E. VANDEN-EIJNDEN, *Commun. Comput. Phys.* **2**, 367 (2007).
- ³⁰ L. MARAGLIANO, A. FISCHER, E. VANDEN-EIJNDEN, and G. CICCOTTI, *J. Chem. Phys.* **125**, 024106 (2006).
- ³¹ G. J. MARTYNA, D. J. TOBIAS, and M. L. KLEIN, *J. Chem. Phys.* **101**, 4177 (1994).
- ³² D. J. TOBIAS, G. M. MARTYNA, and M. L. KLEIN, *J. Phys. Chem.* **97**, 12959 (1993).
- ³³ M. E. TUCKERMAN, J. ALEJANDRE, R. LÓPEZ-RENDÓN, A. L. JOCHIM, and G. J. MARTYNA, *J. Phys. A: Math. Gen.* **39**, 5629 (2006).
- ³⁴ T.-Q. YU, J. ALEJANDRE, R. LÓPEZ-RENDÓN, G. J. MARTYNA, and M. E. TUCKERMAN, *Chem. Phys.* **370**, 294 (2010).
- ³⁵ G. J. MARTYNA, M. E. TUCKERMAN, and M. L. KLEIN, *J. Chem. Phys.* **97**, 2635 (1992).
- ³⁶ Y. LIU and M. E. TUCKERMAN, *J. Chem. Phys.* **112**, 1685 (2000).
- ³⁷ E. VANDEN-EIJNDEN and G. CICCOTTI, *Chem. Phys. Lett.* **429**, 310 (2006).
- ³⁸ G. BUSSI and M. PARRINELLO, *Phys. Rev. E* **75**, 056707 (2007).
- ³⁹ S. MELCHIONNA, *J. Chem. Phys.* **127**, 044108 (2007).
- ⁴⁰ L. MARAGLIANO and E. VANDEN-EIJNDEN, *J. Chem. Phys.* **128**, 184110 (2008).
- ⁴¹ J. G. KIRKWOOD, *J. Chem. Phys.* **3**, 300 (1935).
- ⁴² M. MONTEFERRANTE, S. BONELLA, S. MELONI, and G. CICCOTTI, *Mol. Simulat.* **35**, 1116 (2009).
- ⁴³ W. E, W. REN, and E. VANDEN-EIJNDEN, *Phys. Rev. B* **66**, 052301 (2002).
- ⁴⁴ W. E, W. REN, and E. VANDEN-EIJNDEN, *J. Chem. Phys.* **126**, 164103 (2007).
- ⁴⁵ M. LI, M. DUAN, J. FAN, L. HAN, and S. HUO, *J. Chem. Phys.* **139**, 185101 (2013).
- ⁴⁶ N. DUFF and B. PETERS, *J. Chem. Phys.* **135**, 134101 (2011).

- ⁴⁷ D. QUIGLEY and P. M. RODGER, *J. Chem. Phys.* **128**, 154518 (2008).
- ⁴⁸ D. QUIGLEY and P. M. RODGER, *Mol. Simulat.* **35**, 613 (2009).
- ⁴⁹ W. LECHNER and C. DELLAGO, *J. Chem. Phys.* **129**, 114707 (2008).
- ⁵⁰ A. REINHARDT, J. P. K. DOYE, E. G. NOYA, and C. VEGA, *J. Chem. Phys.* **137**, 194504 (2012).
- ⁵¹ A. B. BELONOSHKO, R. AHUJA, and B. JOHANSSON, *Phys. Rev. Lett.* **87**, 165505 (2001).
- ⁵² A. B. BELONOSHKO, O. LEBACQ, R. AHUJA, and B. JOHANSSON, *J. Chem. Phys.* **117**, 7233 (2002).
- ⁵³ F. SAIJA and S. PRESTIPINO, *Phys. Rev. B* **72**, 024113 (2005).
- ⁵⁴ A. B. BELONOSHKO, *Phys. Rev. B* **78**, 174109 (2008).
- ⁵⁵ M. ROSS and A. K. MCMAHAN, *Phys. Rev. B* **21**, 1658 (1980).
- ⁵⁶ P. LOUBEYRE, *Phys. Rev. B* **37**, 5432 (1988).
- ⁵⁷ M. E. TUCKERMAN, D. A. YARNE, S. O. SAMUELSON, A. L. HUGHES, and G. MARTYNA, *Comp. Phys. Comm.* **128**, 333 (2000).
- ⁵⁸ PINY_MD is available as an open-source code under the Common Public License and capable of performing a wide variety of molecular dynamics and path-integral simulations under different ensembles, using forces calculated from the force fields or from “on the fly” density functional theory electronic structure calculations. It can be downloaded from homepages.nyu.edu/~mt33/PINY_MD/PINY.html.
- ⁵⁹ A. RAHMAN and G. JACUCCI, *Il Nuovo Cimento* **4**, 357 (1984).
- ⁶⁰ D. SHEPPARD, P. XIAO, W. CHEMELEWSKI, D. D. JOHNSON, and G. HENKELMAN, *J. Chem. Phys.* **136**, 074103 (2012).
- ⁶¹ H. CYNN, C. S. YOO, B. BAER, V. IOTA-HERBEI, A. K. MCMAHAN, M. NICOL, and S. CARLSON, *Phys. Rev. Lett.* **86**, 4552 (2001).
- ⁶² <https://code.google.com/p/debyer>.
- ⁶³ P. DEBYE, *Ann. Physik.* **351**, 809 (1915).
- ⁶⁴ Q. MEI and K. LU, *Prog. Mater. Sci.* **52**, 1175 (2007).
- ⁶⁵ G. GRIMVALL, B. MAGYARI-KÖPE, V. OZOLIŅŠ, and K. A. PERSSON, *Rev. Mod. Phys.* **84**, 945 (2012).
- ⁶⁶ R. M. LYNDEN-BELL, J. S. VAN DUJNEVELDT, and D. FRENKEL, *Phys. Rev. Lett.* **80**, 801 (1993).

- ⁶⁷ R. RADHAKRISHNAN and B. L. TROUT, *Phys. Rev. Lett.* **90**, 1786 (2003).
- ⁶⁸ Y. MISHIN, M. J. MEHL, D. A. PAPACONSTANTOPOULOS, A. F. VOTER, and J. D. KRESS, *Phys. Rev. B* **63**, 224106 (2001).
- ⁶⁹ A. B. BELONOSHKO, N. V. SKORODUMOVA, A. ROSENGREN, and B. JOHANSSON, *Physical Review B* **73**, 012201 (2006).
- ⁷⁰ J. R. MORRIS, C. Z. WANG, K. M. HO, and C. T. CHAN, *Phys. Rev. B* **49**, 3109 (1994).
- ⁷¹ L. ZHENG, Q. AN, Y. XIE, Z. SUN, and S. N. LUO, *Journal of Chemical Physics* **127**, 164503 (2007).
- ⁷² S.-N. LUO, A. STRACHAN, and D. C. SWIFT, *J. Chem. Phys.* **120**, 11640 (2004).
- ⁷³ P. G. BOLHUIS, D. CHANDLER, C. DELLAGO, and P. L. GEISSLER, *Ann. Rev. Phys. Chem.* **53**, 291 (2002).
- ⁷⁴ B. PETERS, *J. Chem. Phys.* **125**, 241101 (2006).